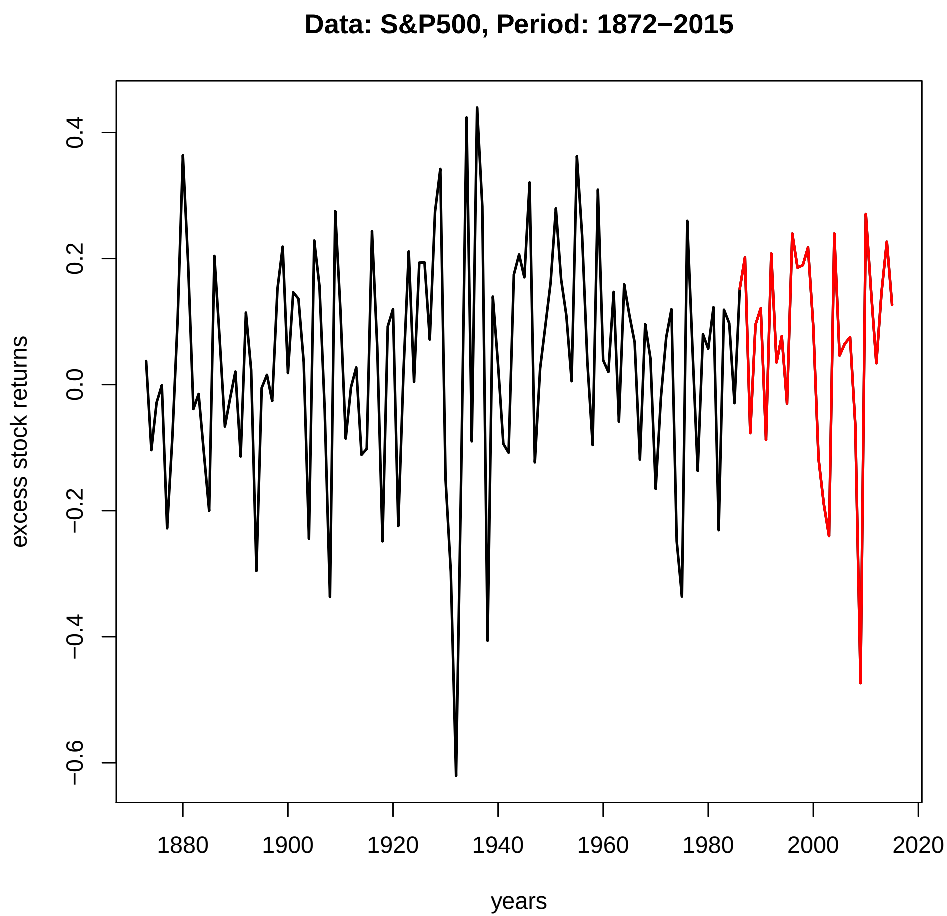
Improved Nonparametric Estimation of the Sharpe Ratio*

* based on joint work with
Stefan Sperlich, Jens Perch Nielsen, and Enno Mammen

Michael Scholz
Department of Economics
University of Graz

CFE–CMStatistics
Sevilla, December 2016


Data: S&P500, Period: 1872–2015
excess stock returns
0.4
0.2
0.0
−0.2
−0.4
−0.6
1880
1900
1920
1940
1960
1980
2000
2020
years

Objectives of the talk:

- Are **equity returns or premiums** predictable? Until mid-1980's: Predictability would contradict the **efficient markets paradigm**.
- Empirical research in the late 20th century and recent progress in asset pricing theory suggest that excess returns are **predictable**.
- We take the **long-term actuarial view** and base our predictions on **annual data** of the S&P500 from 1872 through 2015 on a one year horizon.
- **Our interests**:
  - Actuarial models of **long-term saving** and potential **econometric improvements** to such models.
  - **Market timing/compare assets** based on the Sharpe ratio (SR)

$$SR_t = \frac{\mathbb{E}(Y_t|\boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1})}{\sqrt{Var(Y_t|\boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1})}}$$

Motivation III

- Not many historical years in our records and **data sparsity** is an important issue.
- **Bias** might be of **great importance** when predicting yearly data. Classical trade-off of variance and bias depends on the horizon/frequency.
- **Advocate** for non- and semi-parametric methods in financial applications:
  - **Powerful data-analytic tools**: Local-linear kernel smoothing and wild bootstrap.
  - With **suitable modifications** those techniques can perform well in different economic fields.
  - Include **prior knowledge** in the statistical modelling process for **bias reduction** and to **avoid the curse of dimensionality** and other problems.

Overview:

- The **prediction framework** and the **Sharpe ratio**
- A measure for the **quality of prediction**: The validated R^2
- Improved smoothing through **prior knowledge** and estimation of **conditional mean/variance function**
- **Simulation** and **empirical study**
- Outlook and summary

- The **excess stock returns**:

$$S_t = \log\left(\frac{P_t + D_t}{P_{t-1}}\right) - r_{t-1}$$

with **dividends** D_t paid during period t, **stock price** P_t at the end of period t, and **short-term interest rate** $r_t = \log(1 + R_t/100)$ with **discount rate** R_t

- Consider **one-year-ahead predictions** ($T = 1$), but predictions over the **next** $\boldsymbol{T}$ periods are also easily included:

$$Y_t = S_t + \ldots + S_{t+T-1} = \sum_{i=0}^{T-1} S_{t+i}$$

but this would pose **greater statistical challenges**.

- One traditional equation for the **value of a stock** is

$$P_t = \sum_{j=1}^{\infty} (1+\gamma)^{-j}(1+g)^{j-1} D_t$$

with γ discount rate and g growth of dividend yields.

- Price of stocks depends on quantities such as **dividend yield**, **interest rate**, **inflation** (last two highly correlated with almost any relevant discount rate)
- Covariates $\boldsymbol{X}_t$ with predictive power: **dividend-price ratio**, **earnings-price ratio**, **interest rates**, ...
- Consider the model

$$Y_t = m(\boldsymbol{X}_{t-1}) + \nu(\boldsymbol{X}_{t-1})^{1/2}\varepsilon_t$$

where $\mathbb{E}(\varepsilon_t|\boldsymbol{X}_{t-1}) = 0$ and $Var(\varepsilon_t|\boldsymbol{X}_{t-1}) = 1$.

- A way to examine the performance of an investment by adjusting for its risk (**reward-to-variability ratio**)
- It measures the **excess return** (or **risk premium**) per **unit of deviation** in an investment asset or a trading strategy. Sharpe (JPM 1994):

$$SR_t = \frac{E(Y_t)}{\sqrt{Var(Y_t)}}$$

- Use in finance:
  - SR characterizes how well the **return of an asset** compensates the investor for the **risk taken**.
  - When comparing two assets vs. a common benchmark, the one with a **higher SR** provides **better return for the same risk** (the same return for a lower risk)

- In practice: **ex-post** SR used with **realized** rather than **expected returns**

$$SR_{t,simple} = \frac{mean(Y_t)}{sd(Y_t)}$$

- **Easy** to calculate but **depends on length** of observations, includes both **systematic and idiosyncratic risk**
- **Non-normality** of assets, effect of **covariates**, **predictions**?
- In our setting:

$$SR_t(\boldsymbol{x}_{t-1}) = \frac{\mathbb{E}(Y_t|\boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1})}{\sqrt{Var(Y_t|\boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1})}} = \frac{m(\boldsymbol{x}_{t-1})}{\sqrt{\nu(\boldsymbol{x}_{t-1})}}$$

and we get a **two-step** estimator for the **Sharpe ratio** as

$$\widehat{SR}_t = \frac{\hat{m}_t}{\sqrt{\hat{\nu}_t}}$$

Overview:

- The **prediction framework** and the **Sharpe ratio**
- A measure for the **quality of prediction**: The validated R^2
- Improved smoothing through **prior knowledge** and estimation of **conditional mean/variance function**
- **Simulation** and **empirical study**
- Outlook and summary

In detail:

- Consider $Y_t = \mu + \xi_t$ and $Y_t = g(X_t) + \zeta_t$
- μ estimated by the mean $\bar{Y}$ and g by local-linear kernel regression
- Nielsen and Sperlich (IME 2003) define the **validated** R_V^2 as

$$R_V^2 = 1 - \frac{\sum_t \{Y_t - \hat{g}_{-t}\}^2}{\sum_t \{Y_t - \bar{Y}_{-t}\}^2},$$

  where the function g and the simple mean $\bar{Y}$ are predicted at point t without the information contained in t

- A **cross-validation criterion** to rank different models used for both **choice of bandwidth** and **model selection**.

Properties:

- Replacement of **total variation** and **not explained variation** in usual R^2 by its **cross validated** analogs.
- $R_V^2 \in (-\infty, 1]$
- Measures how well a given model and estimation principle predicts compared to another (here: to the CV mean)
- If $R_V^2 < 0$ we cannot predict better than the mean.
- CV punishes **overfitting**, i. e. pretending a functional relationship that is not really there (leads to $R_V^2 < 0$)
- It is a (non-classical) **out-of sample** measure

Overview:

- The **prediction framework** and the **Sharpe ratio**
- A measure for the **quality of prediction**: The validated R^2
- Improved smoothing through **prior knowledge** and estimation of **conditional mean/variance function**
- **Simulation** and **empirical study**
- Outlook and summary

Improved smoothing through prior knowledge I

- Basic idea: **Combined estimator** Glad (ScanJStat 1998)

  Nonparametric estimator **multiplicatively** guided by, for example, parametric model

$$g(x) = g_\theta(x) \cdot \frac{g(x)}{g_\theta(x)}$$

- **Essential fact**:
  - Prior captures characteristics of **shape** of $g(x)$
  - **Correction factor** $g(x)/g_\theta(x)$ is **less variable** than original function $g(x)$
  - Nonparametric estimator gives **better results** with **less bias**
- Include **prior information** in analysis coming from
  - Good economic model
  - (Simple) empirical data analysis or statistical modelling (**shape**, **constructed variables**, **different frequencies**)

Improved smoothing through prior knowledge II

- **Local Problem**: Prior crosses x-axis
  - More robust estimates with suitable **truncation**:
    Clipping absolute value below $\frac{1}{10}$ and above 10
  - **Shift** by a distance c so that new prior strictly greater than zero and does not intersect the x-axis
- **Dimension reduction**: Use possible **overlapping** covariates $\boldsymbol{x}_1, \boldsymbol{x}_2$ for prior and correction factor

$$g(\boldsymbol{x}_1) = (g_\theta(\boldsymbol{x}_2) + c) \cdot \frac{g(\boldsymbol{x}_1)}{g_\theta(\boldsymbol{x}_2) + c}$$

- Scholz et al. (IME 2015) applied the **combined estimator** to annual American data.
- A bootstrap test on the true functional form of the conditional expected returns **confirms predictability**.
- Including prior knowledge shows **notable improvements** in the prediction of excess stock returns compared to linear and fully nonparametric models.
- We will use their **best model** as starting point in our empirical analysis:
  - Prior: linear model with **risk-free** rate as predictive variable
  - Correction factor: fully nonparametric with **earnings by price** and **long term interest**
  - Gives $R_V^2 = 20.9$ (compared to a $R_V^2 = 14.5$ for a fully nonparametric model without prior)

- Four main approaches proposed in the literature: **direct method**, **residual based**, **likelihood-based**, and **difference-sequence method**
- **Direct method** is based on

$$Var(Y_t|\boldsymbol{X}_t = \boldsymbol{x}_t) = \mathbb{E}(Y_t^2|\boldsymbol{X}_t = \boldsymbol{x}_t) - \mathbb{E}(Y_t|\boldsymbol{X}_t = \boldsymbol{x}_t)^2$$

  where both parts are **separately** estimated. Result is **not nonnegative** and **not fully adaptive** to the mean fct. Härdle and Tsybakov (JoE 1997) with $\boldsymbol{X}_t = Y_{t-1}$.
- **Residual based** methods consist of two stages:
  1. estimate $\hat{m}$ and squared residuals $\hat{u}_t^2 = (Y_t - \hat{m}(\boldsymbol{X}_t))^2$
  2. estimate $\hat{v}$ from $\hat{u}_t^2 = v(\boldsymbol{X}_t) + \varepsilon_t$

- Different variants of residual based method (mostly) for 2nd step:
  - Fan and Yao (Biometrika 1998) apply **loc-lin** in both stages. Result is **not nonnegative** but **asymp. fully adaptive** to the unknown mean fct.
  - Ziegelmann (ET 2002) proposes the **local exponential estimator** to ensure **nonnegativity**

$$\sum_t \left(\hat{u}_t^2 - \Psi\left\{\alpha + \beta(X_t - x)\right\}\right)^2 K_h(X_t - x) \quad \Rightarrow \quad Min_{\alpha,\beta}$$

  - Mishra, Su, and Ullah (JBES 2010) propose the use of the **combined estimator** with a parametric guide. They ignore **bias** reduction in 1st step.
  - Xu and Phillips (JBES 2011) use a **re-weighted local constant** estimator (maximize the empirical likelihood s.t. a bias-reducing moment restriction).

- Yu and Jones (JASA 2004) use estimators based on a **localized normal likelihood** (standard **loc-lin** for estimating the mean m and **loc log-lin** for variance ν)

$$-\sum_t \left\{ \frac{(Y_t - m(X_t))^2}{\nu(X_t)} + \log(\nu(X_t)) \right\} K_h(X_t - x)$$

- Wang et al. (AnnStat 2008)
  - analyze the **effect of the mean on variance fct. estimation**
  - compare the performance of the residual-based estimators to a **first-order-difference-based** (**FOD**) **estimator**: loc-lin on

$$D_t^2 = \frac{(Y_t - Y_{t+1})^2}{2}$$

- Residual-based estimators use an **optimal** estimator for mean fct. in

$$\hat{v}(x) = \sum_t w_t(x)(Y_t - \hat{m}(x_t))^2,$$

works well if in $\hat{m}$ the **bias is negligible**. Bias **cannot be further reduced** in 2nd stage. (FOD: crude estimator $\hat{m}(x_t) = Y_{t+1}$)

- Our strategy: Use **combined estimator with simple linear prior** in both stages
- Reasons:
  - FOD was not convincingly performing in **small samples**
  - We know that the mean fct. is rather **smooth**
  - But **bias reduction** is key due to **sparsity**
  - We cannot compare FOD and residual-based results in terms of R_V^2
  - Maybe FOD together with combined estimator?

Overview:

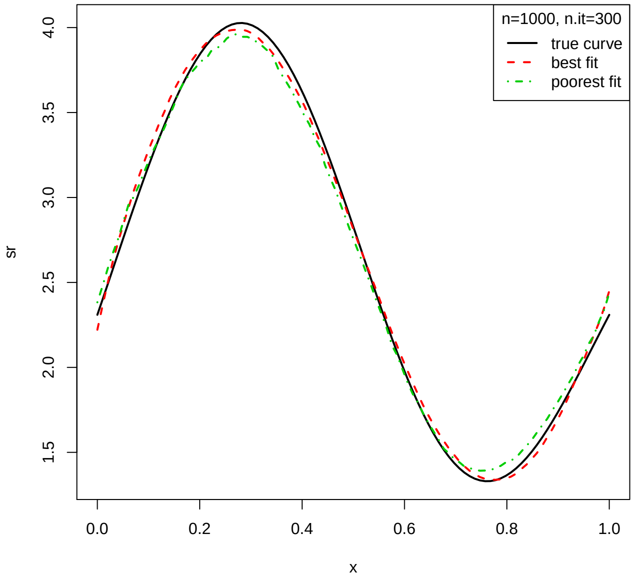
- The **prediction framework** and the **Sharpe ratio**
- A measure for the **quality of prediction**: The validated R^2
- Improved smoothing through **prior knowledge** and estimation of **conditional mean/variance function**
- **Simulation** and **empirical study**
- Outlook and summary

- Which procedure gives reasonable results for estimation of Sharpe ratio?
- Simulate mean and variance function on $x \sim U[0,1]$ as

$$m(x) = 2 + \sin(a\pi x), \quad v(x) = x^2 - x + 0.75 \quad \text{and} \quad sr(x) = m(x)/v(x)^{1/2}$$

- **Best model** in terms of **cross-validated mean square error** (0.020) uses **combined estimator** for mean with a poly3 prior and **combined estimator** for variance with an exponential prior
- **Poorest model** in terms of cross-validated mean square error (0.054) uses **combined estimator** for mean with a exponential prior and **FOD estimator** for variance without prior

- averages of best and poorest model over 300 independent estimates ($n = 1000$)



- **Annual American Data**: Updated and revised version of Robert Shiller's dataset - Chapter 26 in Market volatility (1989)

Tabelle: US market data (1872-2015).

	Max	Min	Mean	Sd
Excess Stock Returns	0.44	-0.62	0.04	0.18
Dividend by Price	0.09	0.01	0.04	0.02
Earnings by Price	0.17	0.02	0.08	0.03
Short-term Interest Rate	17.63	0.19	4.61	2.84
Long-term Interest Rate	14.59	1.91	4.60	2.26
Inflation	0.17	-0.19	0.02	0.06
Spread	3.27	-5.06	-0.01	1.58

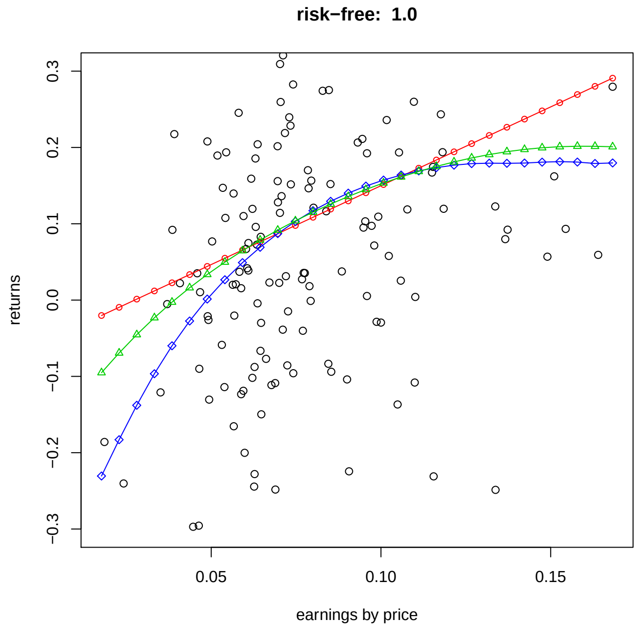
- **estimation of the mean fct.**:

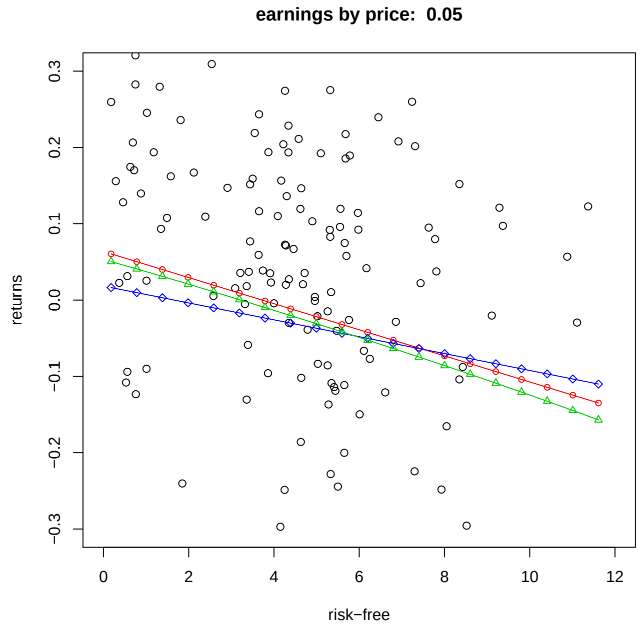
$$m(\boldsymbol{x}_1) = (m_\theta(\boldsymbol{x}_2) + c) \cdot \frac{m(\boldsymbol{x}_1)}{m_\theta(\boldsymbol{x}_2) + c}$$

Tabelle: Predictive power (in percent)

		prior						no prior
		S	d	e	r	L	inf	
corr.	e	8.5	8.4	9.1	16.5	10.9	11.2	11.5
fac-	e, L	9.8	12.8	14.4	20.9	13.0	13.2	14.5
tor	e, r	10.9	9.0	12.1	11.9	10.3	14.0	14.7

- Our choice: linear model with **risk-free** as prior and fully nonparametric with **earnings by price** and **long-term interest rate** for correction factor.


risk−free: 1.0
returns
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
0.05
0.10
0.15
earnings by price


earnings by price: 0.05
returns
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
0
2
4
6
8
10
12
risk−free

- **estimation of the variance fct.**:

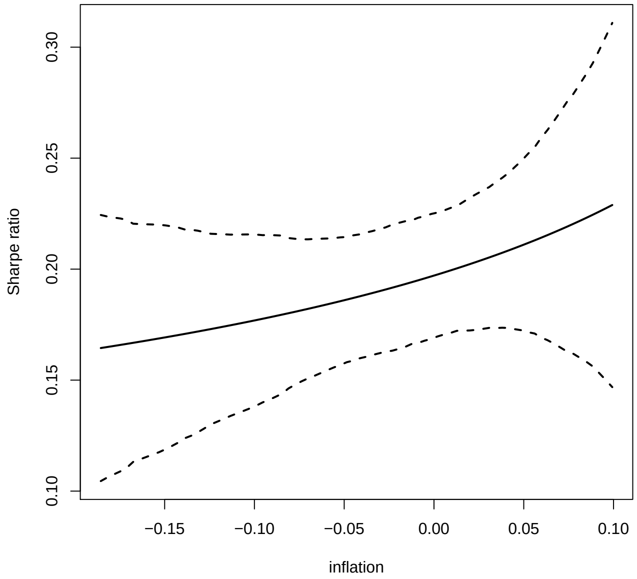
$$\nu(\boldsymbol{x}_1) = (\nu_\theta(\boldsymbol{x}_2) + c) \cdot \frac{\nu(\boldsymbol{x}_1)}{\nu_\theta(\boldsymbol{x}_2) + c}$$

Tabelle: Predictive power (in percent)

		prior in mean&var		prior in mean	no prior
		R^2_V	prior	R^2_V	R^2_V
corr.	S	7.4	L	1.7	0.5
fac-	$S, spread$	5.9	L	1.1	-0.5
tor	$inf, spread$	10.7	L, inf	5.4	-1.2

- In all models **long-term interest** or **long-term interest** and **inflations** as prior
- Best model without any prior: $R^2_V = 1.8$ with **spread**

Estimator of **Sharpe ratio** evaluated at mean values for other covariates than **inflation** with confidence intervals based on **wild bootstrap**



Overview:

- The **prediction framework** and the **Sharpe ratio**
- A measure for the **quality of prediction**: The validated R^2
- Improved smoothing through **prior knowledge** and estimation of **conditional mean/variance function**
- **Simulation** and **empirical study**
- Outlook and summary

- We propose an estimator for the Sharpe ratio as a two stage estimator of conditional mean and variance function using a **combined estimator with parametric priors**.
- Include **prior knowledge** in the statistical modelling process. Improve this way smoothing due to **bias reduction**.
- We provide an **ex-ante measure of asset performance** incorporating the risk taken.
- Could be used for **market timing**. Out-of-sample performance?
- SR appropriate measure during **declining markets**?

Thank you for your attention!